Naval Research Laboratory

Washington, DC 20375-5320

NRL/FR/5550--02-10,047

Evaluation and Improvement of a Speaker Verification System in Military Environments

DAVID A. HEIDE

Transmission Technology Branch
Information Technology Division

December 26, 2002

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
26-12-2002	Formal	Continuing; 15 May 2001-15 Aug. 2002

4. TITLE AND SUBTITLE

Evaluation and Improvement of a Speaker Verification System in Military Environments

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
33904N, 61153N

6. AUTHOR(S)

David A. Heide

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/FR/5550--02-10,047

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Commander
Space and Naval Warfare Systems Command
4301 Pacific Highway
San Diego, CA 92110-3127

10. SPONSOR / MONITOR'S ACRONYM(S)

SPAWAR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The U.S. Navy is constantly looking for ways to improve security. One of the possible solutions proposed has been the use of biometrics, defined by the Biometric Consortium as "automated methods of recognizing a person based on a physiological or behavioral characteristic." One of the least costly, most convenient, and least invasive methods of biometrics is speaker verification.

The Naval Research Laboratory's Voice Systems Section has undergone an extensive study of a commercial speaker verification system in adverse military noise and voice encoding environments. Additional testing was conducted to study the amount of improvement that could be achieved using a noise canceling preprocessor.

This report documents the performance results in 10 different military noise environments, six different voice encoding algorithms, and all combinations of the two, with and without noise cancellation. Results show that speaker verification definitely shows promise under some conditions. In many cases, a significant improvement in performance was achieved by using a noise cancellation preprocessor.

While more testing needs to be done to gauge the level of performance under live conditions, this report shows system integrators where speaker verification could possibly be used.

15. SUBJECT TERMS

Biometrics
Noise cancellation
Speaker verification
Vocoding

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					David Heide
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	23	**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified			202-404-7107

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

CONTENTS

EVALUATION AND IMPROVEMENT OF A SPEAKER VERIFICATION SYSTEM IN MILITARY ENVIRONMENTS

INTRODUCTION

The Naval Research Lab (NRL) is constantly looking for ways to improve security for access to the U.S. Navy's computers, voice networks, radio circuits, and controlled spaces. One possible way to accomplish this task is through the use of biometrics. Biometrics has been defined by the Biometric Consortium as "automated methods of recognizing a person based on a physiological or behavioral characteristic" [1]. One of the most convenient, least costly, and least invasive methods of biometrics available is speaker verification, that is, the identification of a person based on their voice.

In 1995, when speaker verification was still in its infancy in terms of commercial products, the NRL Voice Systems Section undertook a relatively limited study of a commercial speaker verification system, the ITT SpeakerKey™ speaker verification system, in four military noise environments. Results showed that while the system performed well in an office-type environment, performance suffered when subjected to some typical military noise environments [2].

Recently it was decided to test the current version of this commercial speaker verification system. Goals of this test were to

- measure the level of progress in the speaker verification algorithm performance in the past seven years,
- test the system in many more noise environments,
- test the system in a variety of different radio channel and telephone circuit conditions for remote verification applications,
- improve the verification performance in the harshest noise environments by using a noise canceling preprocessor, and
- provide a comprehensive set of results for system designers to decide where speaker verification might meet their requirements.

This report documents this comprehensive effort of testing the speaker verification system in more than a hundred different conditions. The main sections of this report are outlined below.

- **Background –** The report first gives some background into what biometrics are, and why their use can potentially be very advantageous to the military. Then it discusses the more specific advantages of speaker verification.

Manuscript approved September 5, 2002.

- **Testing Conditions –** This section outlines the testing conditions for this study. Each of the military noise environments is described and spectrograms showing the frequency characteristics of each noise environment are presented. For remote verification applications, each of the voice encoders tested is then specified. Lastly, the speaker verification database and the noise cancellation preprocessor are introduced.

- **Test Results –** This section of the report documents the test results for the speaker verification system under all of the testing conditions, shows the level of improvement achieved with the addition of the noise canceling preprocessor, and compares the results with the 1995 version of the speaker verification system.

- **Conclusions and Recommendations –** Finally, some conclusions and recommendations are given.

BACKGROUND

Applications of Biometrics

Many applications of biometrics present significant advantages over the status quo. Depending on the security requirements, biometrics may not be able to replace existing security measures, but they may significantly enhance them as an additional security layer. Some of the possible applications of biometrics in a military environment include:

- *Remote verification of participants in a conference call.* Currently, many conference calls are only regulated with a simple PIN, if at all. With biometrics, conference participants' identities could be automatically distributed to all members of the conference call.

- *Remote verification of a commander giving orders*. Biometrics can also be used in situations where a commander is giving an order remotely. In many situations, the person receiving the order does not personally know the commander giving the order. Biometrics can help verify that the person giving the order is the actual person authorized to do so.

- *Remote verification of a soldier reporting in from the field*. Similar to the previous application, the verification could be turned around, so that the commander is the one who is verifying who is reporting status information from the battlefield. Figure 1 shows this two-way verification scenario, where a system could verify the commander's identity to the soldier and vice versa.

- *Access control to a controlled space or computer.* As with conference calls, access to many controlled spaces and computers are controlled with a simple PIN. Biometrics can give a much higher level of security because they cannot be given away to anyone like a PIN can.

- *Remote verification of a soldier's identity during search and rescue operations.* During search and rescue operations, the rescuers would like to first make sure they are not being called in to a trap whereby the enemy makes a false distress call with a confiscated radio. As Fig. 2 shows, a downed pilot or stranded soldier calling for rescue could first be verified by voice before the helicopter gets within range.

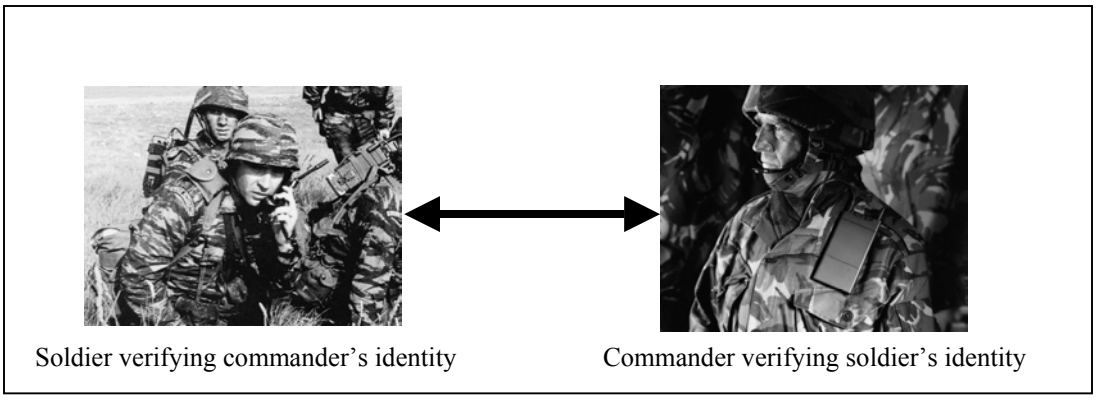

Fig. 1 — This figure shows an application where both parties need verification. The soldier needs to know who is giving him the order. The commander needs to verify who is reporting in from the battlefield.

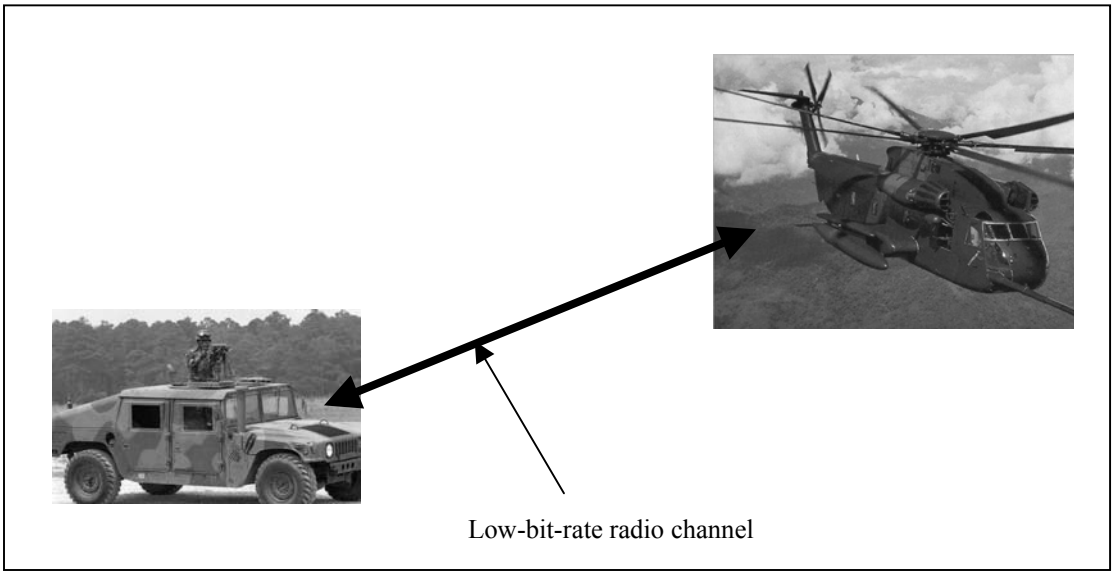

Fig. 2 — This figure shows how a stranded soldier's voice could be first verified before rescue. This verification could help to avoid the possibility of enemy forces setting a trap by calling in a bogus distress call.

Advantages of Speaker Verification

In addition to the advantages of biometrics in general, speaker verification also has its specific advantages.

- *Speaker verification allows for remote verification over existing voice links* – The military has a large amount of legacy communication equipment that only supports voice communication. In this case, the only possible biometric solution for remote verification applications would be speaker verification.

- *Cost* – Even for those applications where communication channels would support data transfer necessary for other biometrics, the input device equipment necessary for speaker verification is already in place. Telephones, tactical radios, and computers already have all the microphones

necessary for voice. With speaker verification, it is not necessary to add fingerprint, hand, or iris scanners at all of the locations because they are already equipped with inexpensive microphones.

- *Speaking is natural and noninvasive* — While some people have expressed fears about having their fingerprint scanned, very few people feel threatened by speaking their password. Because this comes naturally to most people, it generally needs very little training.

TESTING CONDITIONS

Scenarios Tested

The main focus of this study was to judge the performance of a speaker verification system in the following four main scenarios in which the military may use the system.

- *Scenario 1: Quiet environment/uncompressed speech:* This scenario is the baseline environment in which a user would be in a relatively quiet environment with a speaker verification system performing the verification locally. By doing the verification locally, speech is at its highest quality since it is does not have to be encoded over a communication channel. This test gives our baseline, best-case performance of the system.

- *Scenario 2: Noisy environment/uncompressed speech:* This scenario only tests the effect of military noise on the system's performance. Nine military platforms were tested, including military aircraft, Navy ships, and Army personnel carriers. As in scenario 1, all verification would be done locally using high-quality uncompressed speech, but now with military noise corrupting the speech.

- *Scenario 3: Quiet environment/compressed speech:* This scenario involves a user being in a quiet environment, but the verification is done remotely over a communication link. Therefore, the speech would have to be encoded and sent over a communication channel prior to verification. Five possible voice compression algorithms were tested to cover the range of possible communication links that would be encountered by the U.S. military. They ranged from high-quality 32.0 kilobits per second (kb/s) Adaptive Differential Pulse Code Modulation (ADPCM) implemented on the Secure Terminal Equipment (STE) all the way down to 2.4 kb/s Linear Predictive Coder (LPC) found on the Advanced Narrowband Digital Voice Terminal (ANDVT).

- *Scenario 4: Noisy environment/compressed speech:* This scenario combines scenarios 2 and 3 in that the speech is both corrupted by military noise and compressed. For example, a member of the Army on a noisy personnel carrier could be reporting in over a narrowband communication link. This is obviously one of our worst-case scenarios as far as speaker verification systems are concerned. All nine military noise environments were tested with all five voice-compression methods.

Military Noise Environments Tested

In addition to the quiet environment, recordings were made from nine different military platforms to cover a wide variety of possible scenarios of land, sea, and air. Each platform is described and followed by a 10-s spectrogram showing its frequency characteristics.

- P3-C – The P3-C is a four-propeller turboprop aircraft. The spectrogram in Fig. 3 shows two strong areas of stationary noise. One area is below 300 Hz and the other shows a resonance at 3600 Hz.

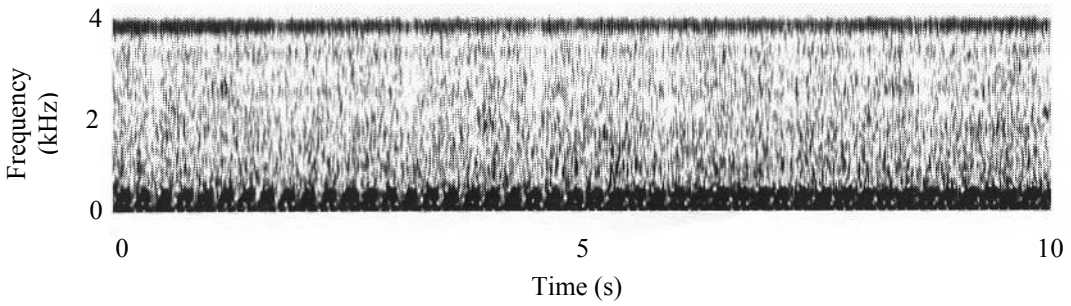

Fig. 3 — P3-C spectrogram

- E3-A – The E3-A is a modified Boeing 707 jet aircraft for use in the Airborne Warning And Control System (AWACS). The noise is much the same as one would hear in older generation commercial aircraft. The spectrogram is shown in Fig. 4.

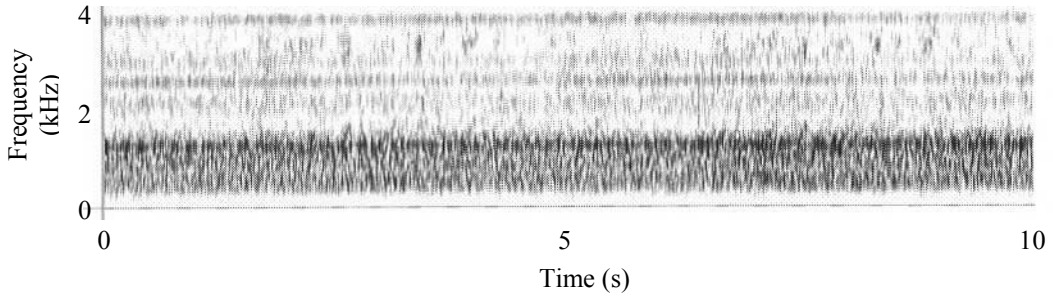

Fig. 4 — E3-A spectrogram

- E2-C – The E2-C is a two-propeller turboprop aircraft. While it is a turboprop aircraft like the P3-C, it is a much smaller aircraft that is much less insulated from the noise. The spectrogram is shown in Fig. 5.

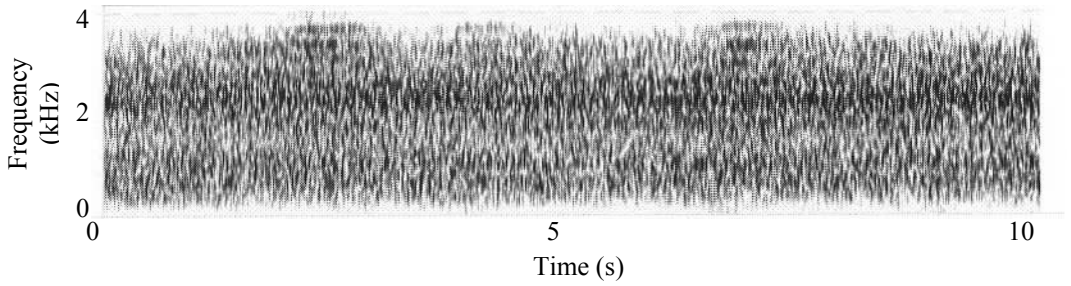

Fig. 5 — E2-C spectrogram

- F15 – The F15 is a jet fighter aircraft. While the noise is harsh, it is somewhat moderated by the oxygen mask of the pilot. The spectrogram is shown in Fig. 6. Notice how at approximately the 4-s point, the pilot accelerates and the noise becomes more intense.

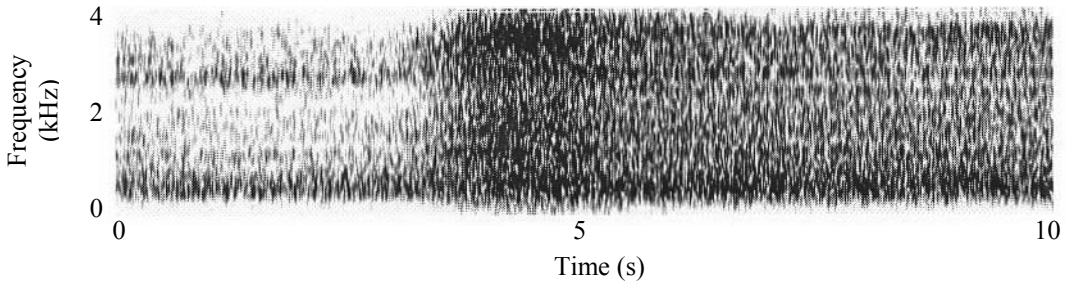

Fig. 6 — F15 spectrogram

- RH-53 – The RH-53 is a military helicopter. The noise is extremely harsh with very strong resonances at 1350 and 2700 Hz. The spectrogram is shown in Fig. 7.

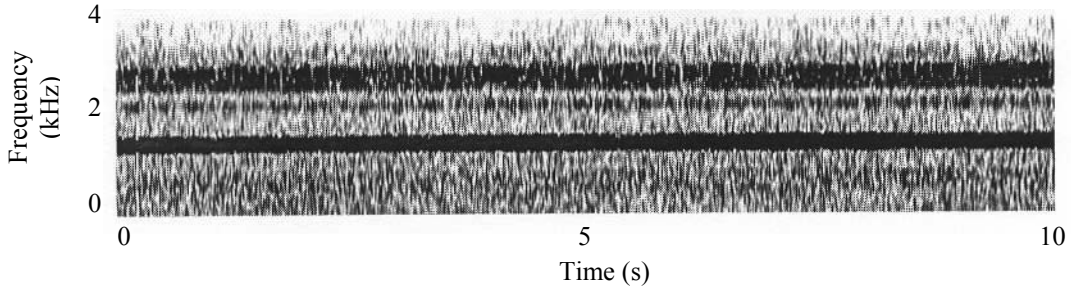

Fig. 7 — RH-53 spectrogram

- Destroyer – This recording is from a very old generation destroyer with very severe noise conditions. This is definitely one of the more difficult ship scenarios tested. The spectrogram is shown in Fig. 8.

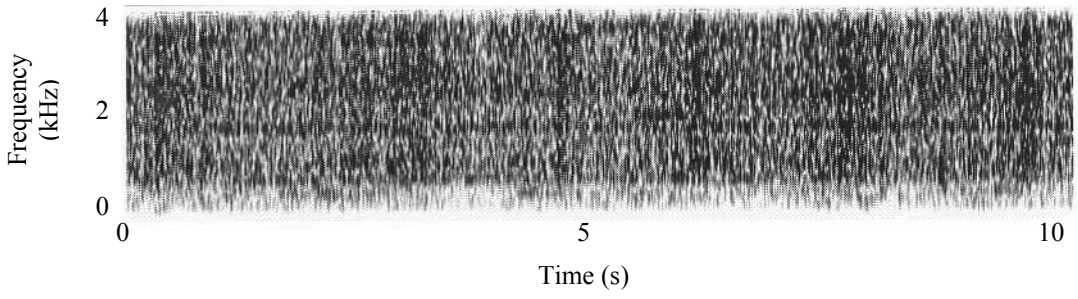

Fig. 8 — Destroyer spectrogram

- USS *George Washington* (GW) – The USS *GW* is a modern aircraft carrier. The noise was recorded in the combat direction center. While it is generally a much less severe environment than the destroyer, note that the spectrogram shows a phone ringing, background speech, and a whistle calling general quarters in only a 10-s time segment. The spectrogram is shown in Fig. 9.

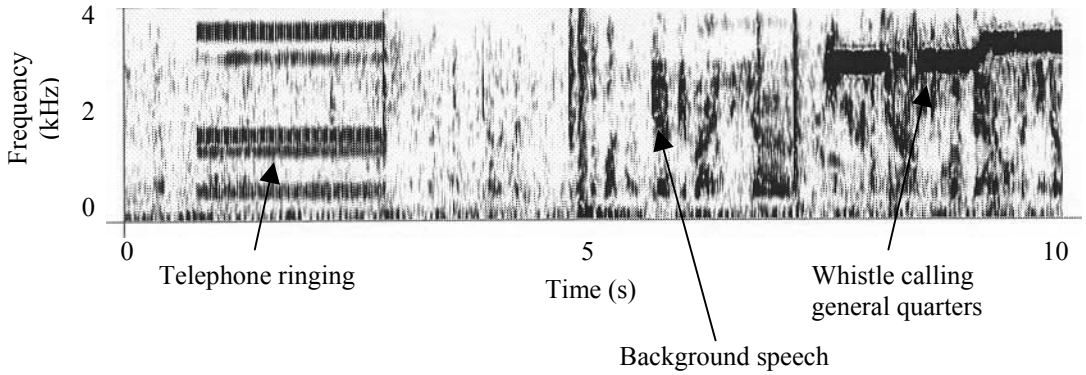

Fig. 9 — USS *George Washington* spectrogram

- M2 Bradley – The M2 is an infantry fighting vehicle. The noise from the M2 is extremely severe. The spectrogram is shown in Fig. 10.

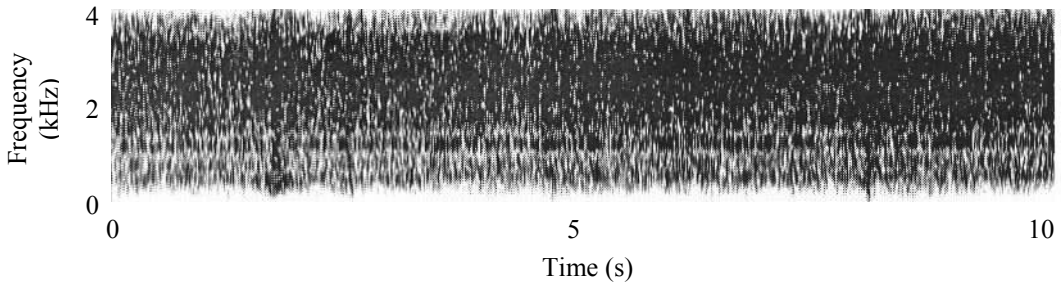

Fig. 10 — M2 Bradley spectrogram

- HMMWV – The HMMWV (High Mobility Multipurpose Wheeled Vehicle) is an Army personnel carrier. The noise from the HMMWV is also extremely harsh. The spectrogram is shown in Fig. 11.

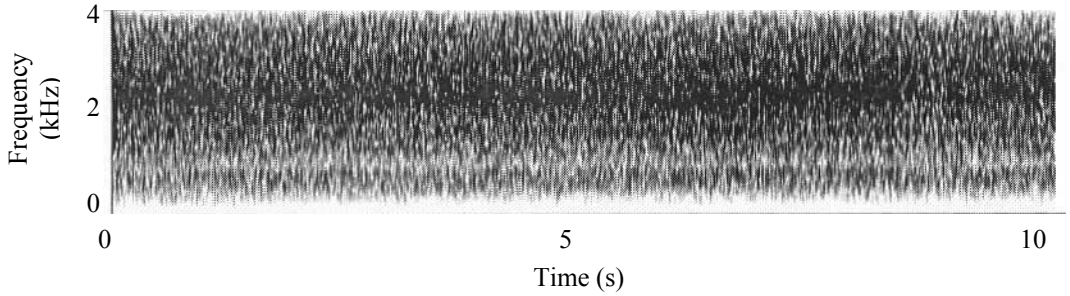

Fig. 11 — HMMWV spectrogram

Voice Encoders Tested

As stated above, one of the most significant advantages of speaker verification is its ability to verify someone *remotely* using existing voice communication equipment. Depending on the equipment used, the quality of this voice link can vary widely. In addition to uncompressed Pulse Code Modulation (PCM), five different voice encoders were tested to simulate the wide variety of communication links that may be used. They are described below.

- PCM (64.0 kb/s) – Pulse Code Modulation (PCM) is just a simple, high-quality speech digitizer found in many office telephones or computers. When we speak of uncompressed speech, we are referring to PCM.

- ADPCM (32.0 kb/s) – Adaptive Differential PCM (ADPCM) is also a very high-quality speech encoder. It is increasingly being used in place of PCM to cut the data rate in half in equipment such as the STE.

- CVSD (16.0 kb/s) – Continuously Variable Slope Delta (CVSD) is a very old generation speech encoder found in many tactical radios such as the Single Channel Ground Airborne Radio (SINCGAR). Speech quality encoded with CVSD is significantly lower than that of ADPCM or PCM.

- CELP (4.8 kb/s) – Codebook Excited Linear Predictor (CELP) is a very complex speech encoder that is found in secure telephones such as the Secure Telephone Unit, Third Generation (STU-III). It gives relatively good speech quality considering its 4.8 kb/s data rate.

- MELP (2.4 kb/s) – Mixed Excitation Linear Predictor (MELP) is a relatively new low bit rate algorithm. While the quality is surely not as high as PCM or ADPCM, it will be used because of its low bit rate and acceptable performance.

- LPC (2.4 kb/s) – Linear Predictive Coder (LPC) is an old generation encoder that MELP was designed to replace. It is still used on many current tactical radios such as the ANDVT and also on the STU-III.

Number of Enrollments Tested

For all biometric devices, users are first enrolled into the system in what is known as an enrollment session. The system takes data, analyzes it, and generates a user's template that defines that particular individual to the system. For some systems, performance can be improved by having multiple enrollment sessions and averaging the results to get a more robust template.

In our testing, we tested many of the first three scenarios (described in the Testing Conditions section) with both one and three enrollment sessions to note how much improvement in performance the three enrollment sessions would give. Scenario 4 (combined noise and voice encoding) was only tested with one enrollment session.

The feasibility of asking people to enroll three separate times certainly varies by application, but perhaps these results could give an idea of how much improvement could be achieved if a system were set to automatically update templates a small amount every time an individual was successfully verified.

Speaker Database

While it is much more desirable to always conduct live testing at each military platform, it is not feasible to do so at ten different platforms (including quiet office) with six different voice encoders (including uncompressed PCM) with a sufficient number of participants. Therefore, all testing for this report was performed using the YOHO speaker database with the appropriate levels of noise added into the reference speech. Then the speech was compressed with the applicable voice encoder. With this setup, we could easily vary any of the testing conditions without needing to bring in the participants each time.

As stated above, all testing was done with the YOHO database. It consists of 138 speakers reading combination lock phrases (24-81-54, 37-39-42, etc.). The speakers were mostly from the New York / New Jersey area, but some non-native English speakers are included. This database is available through the Linguistic Data Consortium at the University of Pennsylvania. The following details about the database are taken from the documentation files supplied with the CD-ROM [3] and Joseph Campbell's report in ICASSP-95 [4].

- 138 speakers (106 males, 32 females)*
- Speech data sampled over a 3-month time frame in an office type environment
- 4 enrollment sessions per speaker with 24 phrases per session
- 10 verification sessions per speaker with 4 phrases per session
- 8 kHz sampling rate with 3.8 kHz analog bandwidth
- 1.2 GB of data when uncompressed

While a successful database test certainly does not guarantee a successful test in the field, it can help to eliminate those conditions where speaker verification may not be the right choice.

Mismatched Enrollment and Verification Conditions

Typically for best performance of speaker verification systems, it is best to match the conditions of the enrollment sessions and the verification sessions. However, in this testing, the much more realistic and appropriate scenario was that of users being enrolled in a quiet environment using high quality uncompressed speech. It is only in the verification session that speakers would be subjected to high noise environments and encoded speech over remote links. While it certainly is possible to force users to enroll over a narrowband communication link inside of a noisy flying P3, for example, it was assumed that they have much more pressing needs at the time. In addition, one would not want to force the users to have multiple enrollment templates for all of the different conditions that they may encounter. *Therefore, in all testing results that follow, the enrollment template is from quiet, uncompressed speech and only the verification sessions are noisy or compressed or both.*

Noise Cancellation Algorithm

Because of our study in 1995, we knew that military noise can seriously affect speaker verification systems. Now that we were making the testing more difficult with the addition of low-bit-rate voice encoders, we knew that a good noise cancellation algorithm may be needed. The algorithm we chose was developed specifically for low-bit-rate voice encoders and was developed jointly by AT&T Labs and the

* Counter to the YOHO readme files.

National Security Agency. The References section lists a number of papers describing this algorithm and voice encoders in which it has been used successfully [5-9].

Summary of Testing Conditions

Figure 12 shows a summary of the testing conditions in a three-dimensional graph. All x/y/z combinations of voice encoder/noise environment/noise cancellation have been tested with one enrollment session. In addition, many of the combinations were also tested with three enrollment sessions. Keep in mind that it would be infeasible to do this large number of tests with live testing. This is where the principal importance of having a good speaker verification database pays off. While databases are expensive to obtain initially, they can be of great use in initial testing of a wide variety of conditions.

TEST RESULTS

The test results section has been divided into four main areas.

- *Noisy Environment/Uncompressed PCM Speech (Fig. 13):* This graph shows the effect of the military noise environments on performance without any additional noise cancellation. In addition, the results for the baseline quiet office environment are included for comparison. Results are shown for one and three enrollment sessions.

- *Quiet Environment/Encoded Speech (Fig. 14):* This graph shows the effects of the voice encoders (including uncompressed PCM) on performance. Results are shown for one and three enrollment sessions.

- *Noisy Environment/Encoded Speech (Figs. 15 through 20):* These six graphs show the combined effects of the military noise platforms and the voice encoders. Each graph shows the results for a particular encoder; the results are shown *with and without noise cancellation.* These six graphs combine to show the results for 108 different tests. In these graphs, results are shown for one enrollment session.

- *Comparison of Results of Current Speaker Verification System to Results from 1995 (Fig. 21):* This last graph compares the current results to results obtained in 1995. The four common noise environments tested are the P3-C, E3-A, USS *George Washington*, and the destroyer.

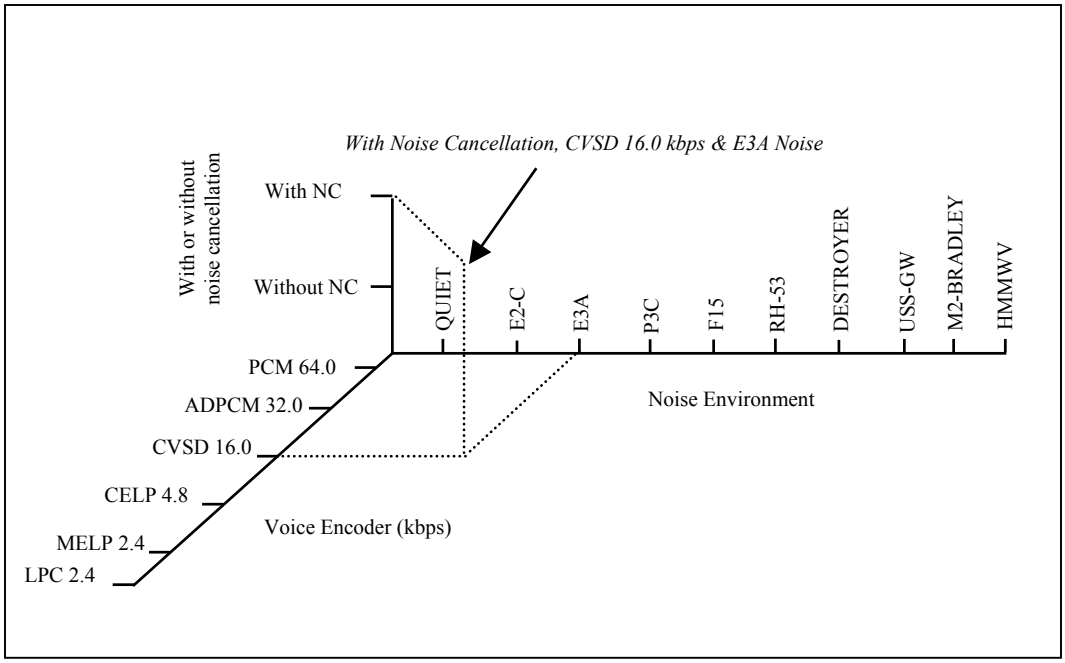

Fig. 12 — This figure shows all of the combinations of Voice Encoder, Noise Environment, and Noise Cancellation that were tested. The arrow shows, for example, the combination with noise cancellation, the 16.0 kbps CVSD voice encoder, and the E3A noise environment.

Quiet and Noisy Environments / Uncompressed Speech

Figure 13 shows how the speaker verification system performed under a variety of noise environments using uncompressed speech. This is the scenario where the verification would be done locally, without the speech being sent over any communication links. In this figure, results are compared using one and three enrollment sessions. No noise cancellation was used.

Note that even without noise cancellation, the P3-C, E3-A, and the USS-GW all have error rates below 1%, and the E2-C and the RH-53 have error rates near 2%. Only in the destroyer, the HMMWV, and the M2-Bradley did error rates significantly rise. While three enrollment sessions helped performance in every case, most platforms experienced less than 0.5% decrease in error rates.

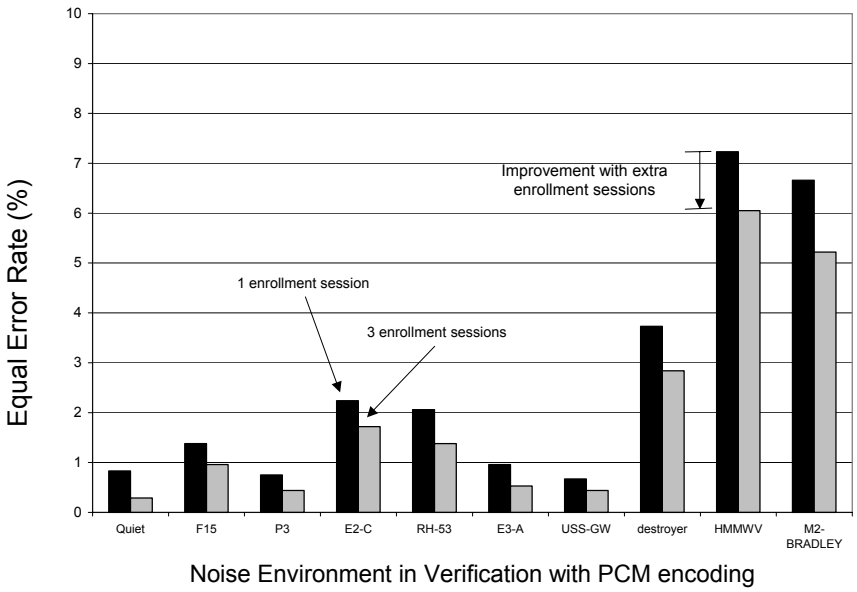

Fig. 13 – Error rates for speaker verification system in noisy environments. This figure shows the performance in ten different noise environments with no speech compression. The side-by-side bars show the difference in performance using one and three enrollment sessions. Note that except for the destroyer, HMMWV, and the M2-Bradley, error rates are all less than approximately 2%.

Quiet Environment / Encoded Speech

Figure 14 shows the results for the remote verification scenario from a relatively quiet environment. In addition to uncompressed PCM, the results from five other voice encoders are shown. Results are shown with one and three enrollment sessions.

The results show that the most modern voice encoders (PCM, ADPCM, CELP, and MELP) performed very well, all with error rates under 1%. Only in the oldest generation voice encoders did performance rise up to 2% and 3 % with LPC and CVSD, respectively, with one enrollment session. Three enrollment sessions again helped performance only marginally, except for the LPC and CVSD, where it reduced the error rates by almost 1%. The next section presents an analysis of the poor performance with the CVSD encoder.

Noisy Environment / Encoded Speech

The next six figures show the combined effect on performance of the noise and the encoding. Each figure shows the results for one voice encoder, with and without noise cancellation. These graphs only show the results for one enrollment session.

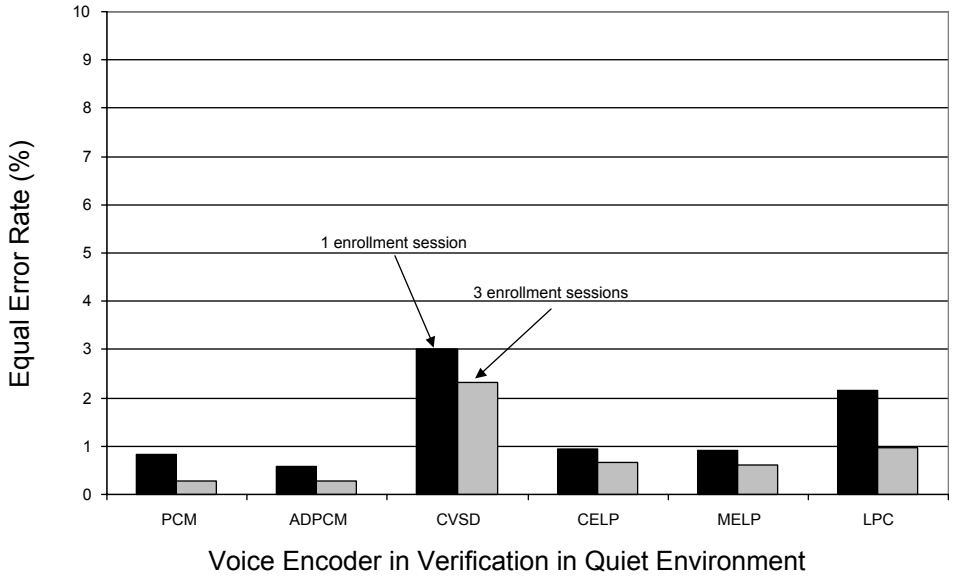

Fig. 14 – Error rates for speaker verification system when speech is encoded (one and three enrollment sessions). This figure shows the performance if a speaker was remotely verified using a low-data-rate voice encoder. Note that only in the older generation voice encoders (CVSD and LPC) did performance suffer.

PCM and ADPCM in Noise Evaluation (Figs. 15 and 16):

Figures 15 and 16 show the results *with and without noise cancellation* for PCM and ADPCM, respectively, in all nine military platform noise environments. The results are very similar because both voice encoders are high data rate and high quality. In general, performance was best in the F15, P3-C, E3-A, and the USS-GW. In those cases, noise cancellation did not improve performance, because the error rates were already mostly below 1%. Noise cancellation tended to help most in the more harsh noise environments, by cutting error rates by approximately 40%.

CVSD in Noise Evaluation (Fig. 17):

Figure 17 shows the results *with and without noise cancellation* for CVSD in all nine military platform noise environments. With CVSD, all error rates have significantly increased. While CVSD is a much lower quality speech encoder than ADPCM, its error rates are even worse than expected, even with the less harsh noise environments and the quiet environment of the previous section.

By conducting additional testing, we found that much of the problem here is with the mismatched conditions of enrollment sessions (PCM encoding) and the verification sessions (CVSD encoding), and not just the noise. By matching the enrollment and verification conditions with CVSD encoding, performance was significantly improved.

To give some sample results, in the quiet environment, the error rate was reduced from 3.0% to 0.9% just by using the CVSD encoder in both the enrollment and verification sessions. In the F15 environment, the error rate was cut in half from 5.5% to 2.7% just by matching the CVSD encoder in enrollment and verification. How realistic it is to require a special CVSD voice template will certainly depend on the application, but this is one special case where it may be worth it.

CELP and MELP in Noise Evaluation (Figs. 18 and 19):

Figures 18 and 19 show the results *with and without noise cancellation* for CELP and MELP, respectively, in all nine military platform noise environments. Results are similar for each encoder. These graphs show where the noise cancellation preprocessor really shows its worth. In many cases, the error rates are cut in half, bringing possibly unacceptable results into a much more satisfactory range. For example, in Fig. 19, the error rate for E2-C in MELP encoding was reduced from 4.6% to 2.2%.

LPC in Noise Evaluation (Fig. 20):

Figure 20 shows the results *with and without noise cancellation* for LPC in all nine military platform noise environments. Because LPC is a very-low-bit-rate and old generation voice encoder, the results are expectedly poor. Noise cancellation tended to help performance significantly, but in many cases the error rates are still too high for most applications. In these cases, it would be much more advisable to try to do the verification locally or with a higher quality voice encoder.

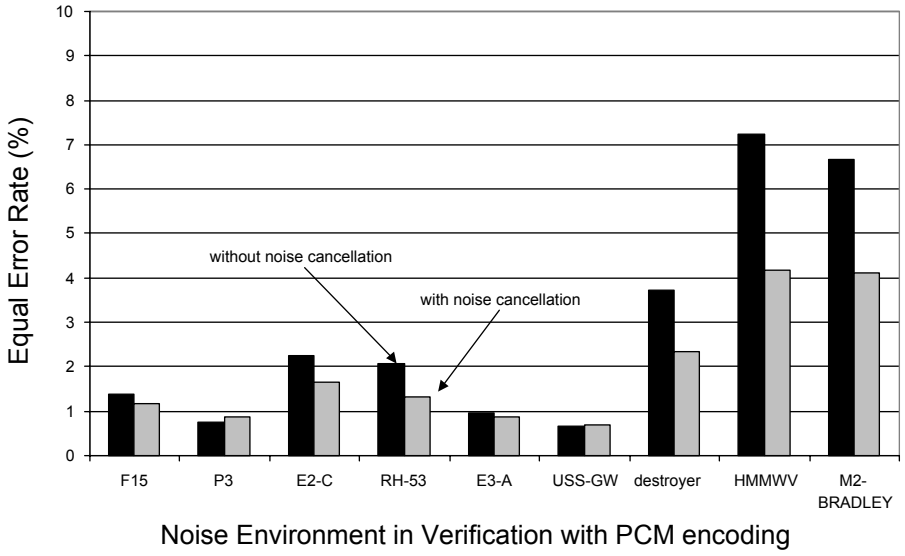

Fig. 15 – Error rates for the speaker verification system in noisy environments when speech is uncompressed (one enrollment session). Only in the harshest environments did error rates rise significantly above 2%. In four cases, it was near or below 1%. The noise cancellation helped the most in the severe environments.

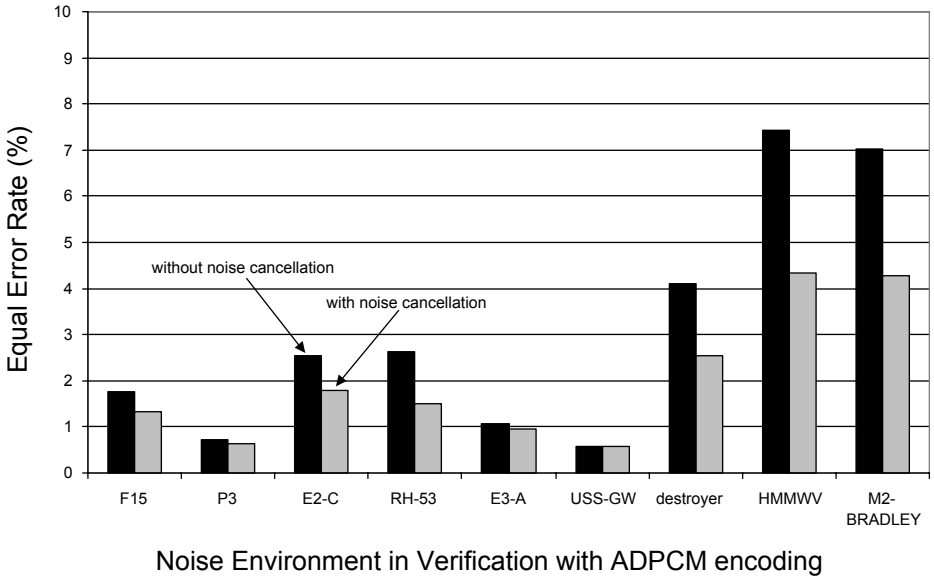

Fig. 16 – Error rates for the speaker verification system in noisy environments when speech is compressed using 32.0 kbps ADPCM (1 enrollment session). Here the results are very similar to those of Fig. 15 because both PCM and ADPCM are very high quality speech encoders.

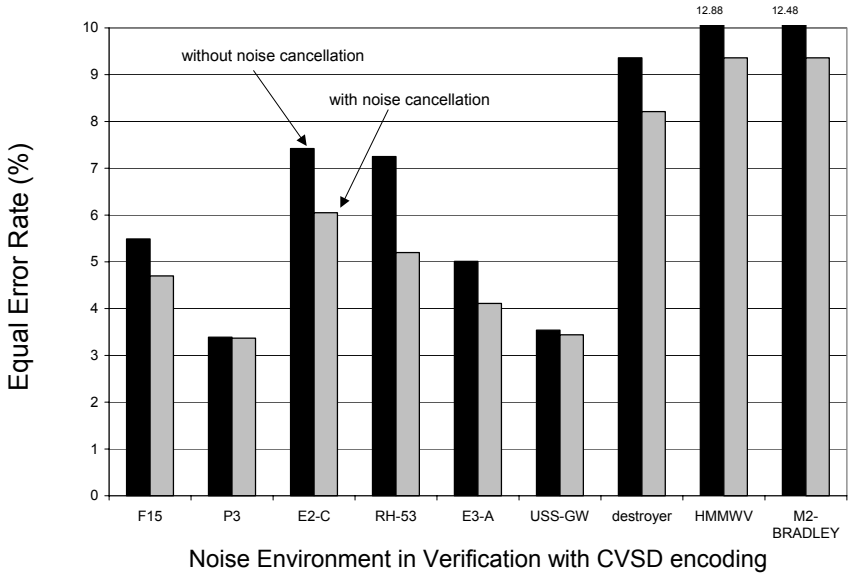

Fig. 17 – Error rates for the speaker verification system in noisy environments when speech is compressed with 16.0 kbps CVSD (one enrollment session). Here the results are significantly worse than one would expect. With further testing, it was learned that much of the problem was in the mismatching of the encoder in the enrollment and verification sessions, and not just from the military noise.

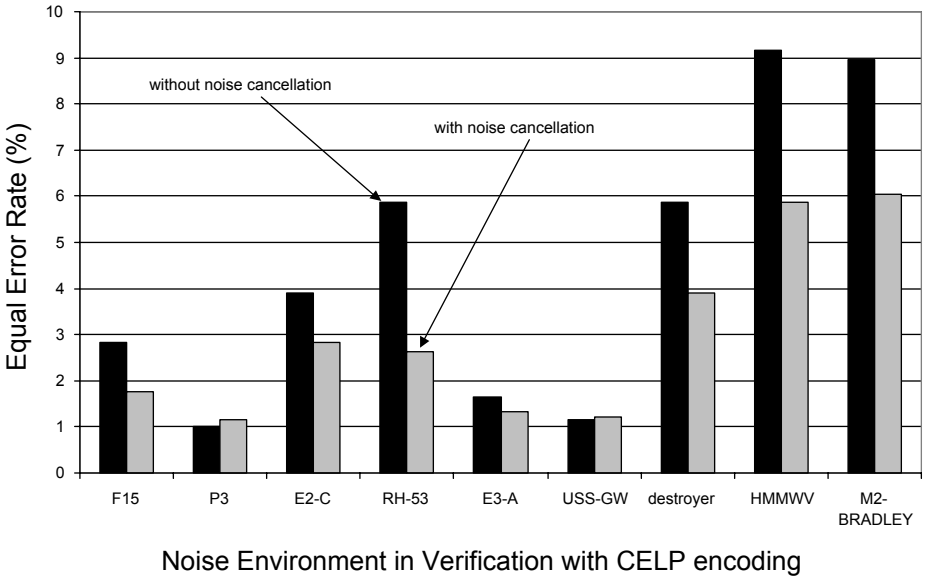

Fig. 18 – Error rate for speaker verification system in noisy environments when speech is compressed with 4.8 kbps CELP (one enrollment session). Here the lower data rate voice encoder starts to hurt performance. While the speaker verification system still performed very well in the less harsh environments (USS-GW, E3-A, P3-C), performance suffered in the others.

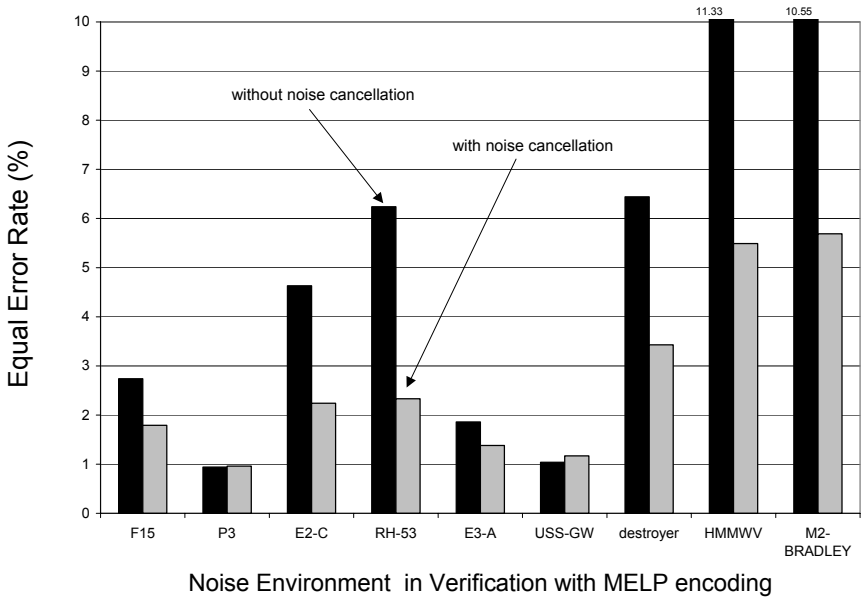

Fig. 19 – Error rate for speaker verification system in noisy environments when speech is compressed with 2.4 kbps MELP (one enrollment session). Results here are very similar to that of the CELP encoder in Fig. 18. The speaker verification system still performs well in the less severe noise environments. Even though noise cancellation helped significantly in many of the other more harsh environments, performance was still poor in these cases.

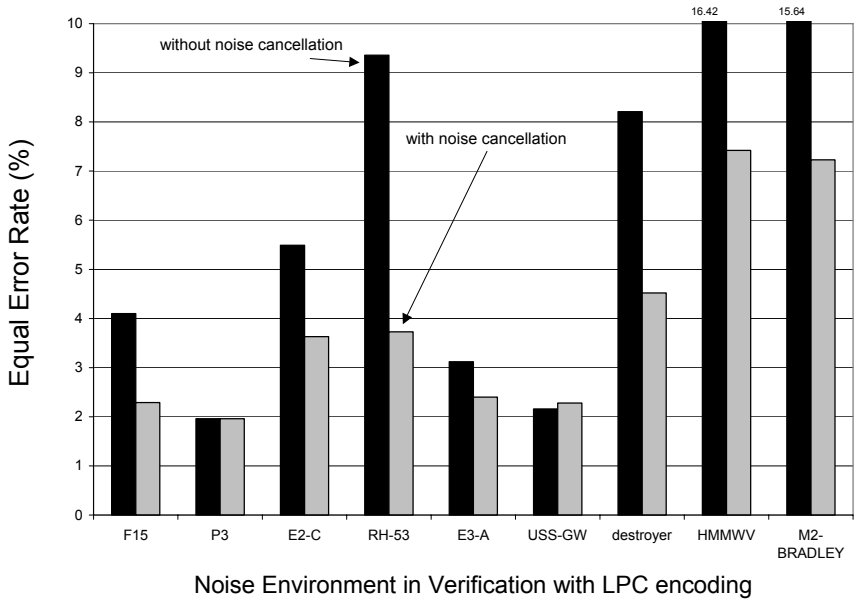

Fig. 20 – Error rate for speaker verification system in noisy environments when speech is compressed with 2.4 kbps LPC (one enrollment session). Results in this figure show that performance is very poor with the old generation LPC voice encoder in all of the severe noise environments.

Comparison of Current Year's Results to Results from 1995 (No Noise Cancellation)

Figure 21 shows the exceptional progress that has been made in the past 7 years in speaker verification in military noise environments. In 1995, only four military noise environments were tested (P3-C, E3-A, USS-GW, and destroyer). In a direct comparison of results from 1995 (uncompressed PCM speech, no noise cancellation, and one enrollment session), error rates were reduced by an average of 70%. What were possibly unacceptable results previously may have improved to an acceptable level.

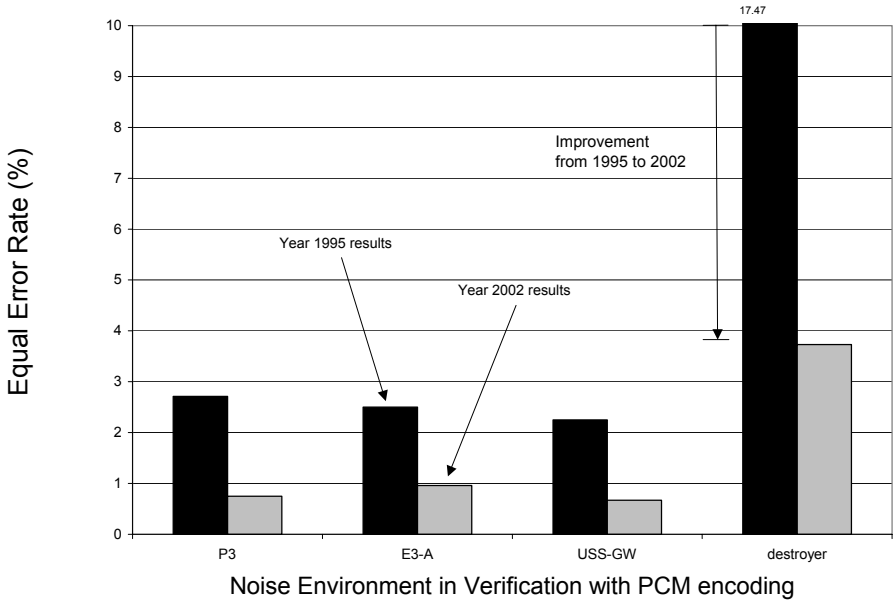

Fig. 21 – Error rate for year 1995 and 2002 speaker verification system in four noisy environments when speech is uncompressed (one enrollment session). This figure shows the dramatic improvement that has been made in speaker verification performance over the past 7 years.

For an even earlier comparison of where speaker recognition was in 1985, the reader is directed to an excellent paper by NRL's Stephanie Everett in ICASSP-85 [10]. She tested one of the earliest prototypes of this speaker recognition system with voice encoders still used in the military today. While the results are not directly comparable because the application and speaker databases were significantly different, it shows how long the Navy has been interested in this problem.

CONCLUSIONS AND RECOMMENDATIONS

- *A tremendous amount of progress in speaker verification performance has been accomplished in the past 7 years.* As shown and described in the Fig. 21, error rates have significantly decreased in military noise environments, even before noise cancellation was used. This amount of progress shows that system integrators must show care before dismissing a particular speaker verification system's performance as unacceptable. By the time the system is finally implemented, later generations of the speaker verification system may have improved significantly. Conversely, system integrators may also want to be cautious to not implement speaker verification in a severe environment where it is not ready. While later generations of the speaker verification system may have been acceptable, the reputation of speaker verification may have already been hurt considerably.

- *For those situations where speaker verification performance shows promise, live testing should be the next step.* Database testing is excellent for judging initial performance over a large number of conditions (10 noise environments vs 6 encoding methods vs noise cancellation), but only live testing can give a final determination of where this system can be deployed. Live testing can also help judge the effects of user stress, microphone variability, throughput rates, etc., to help in that determination.

- *Speaker verification must be implemented on a case-by-case basis:* As noted in all of the results given above, performance varies widely based on noise environment, voice encoding method, and the combination of the two. Therefore, it is imperative that speaker verification is chosen on a case-by-case basis for implementation. The early indications show that speaker verification shows promise in some situations, is marginal in other situations, and is definitely impractical in some other situations. Of course, a platform's security requirements also play a large part in deciding what level of performance is satisfactory.

- *Threshold settings must be determined individually for each application based on noise levels, voice encoding, and security levels desired.* Live testing can give the system administrator a good idea of where to set the accept/reject threshold level based on one's particular security requirements. One of the significant findings from this year's testing was in noting the large range of proper threshold levels over various conditions. The proper threshold setting for a quiet environment could give terrible results in many other conditions and vice versa. As said above, implementing speaker verification must be done on a case-by-case basis, and one generic threshold setting does not carry over to all possible conditions. Keep in mind that all results given in this report are for the equal error rate (the rate at which the false accept rate equals the false reject rate). For high security applications, the system administrator may want to set a more stringent threshold to lower the false accept rate (while raising the false reject rate). For lower security applications, the system administrator may want to do the reverse. All of these questions are best answered with live testing.

- *In the more severe noise environments, noise cancellation can help performance significantly.* In some of the noisy conditions, error rates were reduced by 50%, especially for the low-bit-rate voice encoders. Some of the conditions where the results were unsatisfactory without noise cancellation were improved enough where they might be adequate for some applications.

- *For those very narrowly focused missions where one knows the exact conditions where this system will be used, it may improve performance by trying to match the noise or voice encoding conditions for both enrollment and verification.* As discussed in the test results section, a significant improvement in performance was achieved by matching the CVSD encoder in both the enrollment and verification sessions. However, for most other general cases where a wide variety of conditions can be expected, it is certainly best to just enroll in a quiet environment with no voice encoding for one's user template.

Finally, note that none of these results presented in this report should be interpreted as either an endorsement or rejection of this speaker verification system. In addition to the widely varying noise environments and communication channels, all potential applications have widely varying security requirements that must be taken into account before deciding upon the applicability of implementing this speaker verification system.

ACKNOWLEDGMENTS

The author thanks Vanessa Hallihan and Jim Davies of the Space and Naval Warfare Systems Command (SPAWAR) and David Guerrino, Navy representative to the Biometrics Management Office, for supporting this work. The author also thanks John Collura of the National Security Agency for providing the noise canceling algorithm and Dr. Alan Meyrowitz of NRL for reviewing this report. Finally, the author thanks George Kang, Head of the Voice Systems Section at NRL, for much guidance on this project.

REFERENCES

1. Biometric Consortium Website: http://www.biometrics.org/html/introduction.

2. D.A. Heide, "Evaluation of a Speaker Verification System in the Presence of Noise," International Conference on Signal Processing Applications and Technology (ICSPAT) Boston, MA, October 24-26, 1995, pp. 1974-1978.

3. README Files, YOHO CD-ROM database available from the Linguistic Data Consortium, 441 Williams Hall, University of Pennsylvania, Philadelphia, PA 19104.

4. J.P. Campbell, Jr. "Testing With The YOHO CD-ROM Voice Verification Corpus," IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP), Detroit, MI, May 9-12, 1995, pp. 341-344.

5. J.S. Collura, "Speech Enhancement and Coding in Harsh Acoustic Noise Environments," IEEE Workshop on Speech Coding, Haikko Manor, Finland, 1999, pp. 162-164.

6. R. Martin and R.V. Cox, "New Speech Enhancement Techniques for Low Bit Rate Speech Encoding," Proceedings Speech Coding Workshop, 1999, pp. 165-167.

7. R. Martin, H-G. Kang, and R.V. Cox, "Low Delay Analysis/Synthesis Schemes for Joint Speech Enhancement and Low Bit Rate Speech Coding," Proceedings Eurospeech, Budapest, Hungary, 1999, Vol. 3, pp. 1463-1466.

8. T. Wang, K Koishida, V. Cuperman, A. Gersho, and J.S. Collura, "A 1200 BPS Speech Coder Based on MELP," IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, June 5-9, 2000, pp. 1375-1378.

9. J. Skoglund, R.V. Cox, and J.S. Collura, "A Combined WI and MELP Coder at 5.2 KBPS," IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, June 5-9, 2000, pp. 1387-1390.

10. S.S. Everett, "Automatic Speaker Recognition Using Vocoded Speech," IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP), Tampa, FL, March 26-29, 1985, pp. 383-386.